PRM217

THE USE OF PROPENSITY SCORE MATCHING DOES NOT PROTECT AGAINST REGRESSION ARTIFACTS (REGRESSION TOWARDS THE MEAN)

Caron C[1], Wasser T[2], Eisenberg D[2]

[1]Reading Hospital, West Reading, PA, USA, [2]HealthCore Inc., Wilmington, DE, USA

OBJECTIVES: Propensity Score Matching (PSM) is a common method in many retrospective studies to control for differential treatments. PSM controls for variables where patients are selected for one treatment over another based on aspects of their care that are unknown to the researcher or not a part of the study. This study uses simulated data comparing two cohorts within a population treated for a common psychiatric disorder. Data are analyzed to determine if regression artifacts (RA) are present in the data, uncontrolled by PSM. RA in this context are Type I errors. **METHODS:** Variables commonly used to diagnose patients with Major Depression were simulated: Age, Gender, Ethnicity, Global Assessment of Functioning, Beck Depression and Beck Anxiety scores. Distributions of N=100,000 were simulated for each variable using population values. From these distributions, samples of n=100, n=250 and n=500 were drawn based on typical values that would be seen in a patient with Major Depression. The outcome measure Dependent Variable was the score on the Beck Depression scale, using success of treatment values from 10-15 percent, and correlated with the pretest score using Chomsky's decomposition. PSM was used on a ratio of 1:1. Analysis methods were group and paired t-tests as well as a difference in difference analysis at the end of the study. **RESULTS:** Type I error occurred in each simulation and were correlated with sample size. RA, leading to Type I error were more common at lower sample sizes, in excess of 70%, to a minimum of 54% for n=500. **CONCLUSIONS:** This study demonstrates that RA occur in basic experiments designed to specify treatment effects. Researchers who use PSM methods need to be aware of situations where RA are likely to occur. Standard statistical controls for RA are being tested to see if they correct for RA and Type I error when PSM is used.

PRM218

APPLICATION OF SIMPLE IMPUTATION TECHNIQUES FOR MISSING PAIRWISE CONTRASTS FROM MULTI-ARM TRIALS WHEN USING FREQUENTIST NETWORK META ANALYSIS

Petto H[1], Brnabic A[2], Kadziola Z[1], Belger M[3]

[1]Eli Lilly Regional Operations GmbH, Vienna, Austria, [2]Eli Lilly, Sydney, Australia, [3]Eli Lilly and Company Ltd, Windlesham, UK

OBJECTIVES: When conducting frequentist (fixed effects or random effects) network meta-analysis (NMA), input data is usually required in contrast form. In practice, multiple-arm trials are quite common and results for only the contrast relative to one treatment group are presented. However, some frequentist NMA require all possible pairwise treatment effects and standard errors combinations. While the missing effect sizes can still be directly derived, additional assumptions about covariances are needed to calculate standard errors. **METHODS:** Simple imputation techniques are used for substituting the standard errors of the missing comparisons and this has been applied to both simulated data as well as a real world data example. After imputation data is analyzed using standard frequentist NMA, incorporating multi arm studies by the method described in Rücker (2015). **RESULTS:** We derive simple imputations techniques by (1) assuming independence between contrasts, (2) estimating missing co-variances from the available contrasts in the multi arm trials and (3) from the other two arm studies in the network. Comparable results to networks including all pairwise contrasts can be obtained, especially if only few contrasts are missing in multi arm studies and if variances of the comparisons are not too different. In the first case, even (1) can give acceptable results. If variances differ, but are similar to that from two arm studies then (3) might be preferable over (2). **CONCLUSIONS:** Our results suggests that from a practical point of view, simple imputation techniques might be useful tools for incorporating multi arm trials with incomplete pairwise contrasts into frequentist NMA, although limitations need to be carefully considered. Rücker G: Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods, 2012, 3, 312–324.

PRM219

INDIRECT COMPARISONS IN BENEFIT ASSESSMENT

Kühnast S[1], Schiffner-Rohe J[2], Rahnenführer J[1], Leverkus F[2]

[1]Technical University of Dortmund, Dortmund, Germany, [2]Pfizer Deutschland GmbH, Berlin, Germany

OBJECTIVES: With the Act on the Reform of the Market for Medicinal Products (AMNOG) in Germany, pharmaceutical entrepreneurs must submit a dossier demonstrating additional benefit of a new drug compared to an appropriate comparator. Underlying evidence was planned for registration purposes and therefore often does not meet the appropriate comparator as defined by the Federal Joint Committee (G-BA). For this reason AMNOG allows indirect comparisons (ICs) to assess the extent of additional benefit. This study evaluates the applicability of available IC methods in several situations common to benefit assessment in oncological indications. **METHODS:** An extensive literature search on available statistical methods for performing ICs is performed. Additionally, benefit dossiers containing ICs are analyzed regarding the applied methodology. We use simulation studies to evaluate and compare adjusted (Bucher) and unadjusted methods regarding their properties under different circumstances. **RESULTS:** Adjusted ICs are deemed to be "state of the art". Due to their requirements they are, nevertheless, often not applicable. In most cases reasons are lacking comparability of the trials, e.g. concerning the common comparator, the study population and the study design. Simulations of Hazard Ratios for endpoints overall survival and progression free survival were performed considering various "extents of additional benefit" according to IQWiG criteria. Starting with a setting of identical studies we stepwise modified study population and various attributes in study design. Finally the common comparator was omitted. Discrepancies between ICs and true values are compared graphically and on the basis of statistical measures. **CONCLUSIONS:** ICs imply a set of requirements to be able to derive valid statements. Prerequisites for adjusted ICs are often not met as necessary studies and publications are not available. With respect to the progress of benefit assessment and the subsequent price negotiation it would be helpful having alternatives with acceptable properties in order to estimate the extent of additional benefit.

PRM220

THE USE OF INTERQUARTILE DEVIATION IN ESTABLISHING DELPHI PANEL CONSENSUS: A PRIORITIZATION OF INTRAVENOUS IMMUNOGLOBULIN UTILIZATION

Orange J[1], Lennert B[2], Rane P[3], Eaddy M[2]

[1]Texas Children's Hospital, Houston, TX, USA, [2]Xcenda, Palm Harbor, FL, USA, [3]University of Houston, Houston, TX, USA

OBJECTIVES: To use consensus-building methodologies to prioritize disease states for intravenous immunoglobulin (IVIG) utilization while considering disease severity and alternative therapeutic options. **METHODS:** A 7-member expert panel independently ranked 50 disease states across 2 domains: (1) Disease severity (DS) (1=immediately life-threatening, 2=life-threatening, 3=life-modifying, 4=other) and (2) the perceived efficacy of therapeutic alternatives (TA) (1=none, 2=low, 3=medium, 4=high). An interquartile deviation of $\leq$0.5 was used to determine consensus for disease states within each domain. Disease states reaching consensus across both domains were ranked according to a 4x4 algorithmic scale to establish priority. **RESULTS:** The panel reached consensus on the severity of all diseases states; however, 11 of the 50 disease states did not reach consensus on the availability of alternative therapeutic options. No disease state was designated as being immediately life-threatening without an available alternative therapeutic option (DS1TA1), while 3 disease states (X-linked agammaglobulinemia, common variable immunodeficiency, primary immunodeficiency with absent B-cells) were designated as life-threatening with no therapeutic alternatives (DS2TA1). The priority distribution of disorders based on the algorithm is as follows: DS1TA1=0, DS1TA2=1, DS1TA3=1, DS1TA4=1 DS2TA1=3, DS2TA2=4, DS2TA3=3, DS2TA4=1 DS3TA1=0, DS3TA2=7, DS3TA3=14, DS3TA4=0 DS4TA1=0, DS4TA2=0, DS4TA3=3, DS4TA4=1 **CONCLUSIONS:** The application of interquartile deviation in establishing consensus across two 4-point Likert scales resulted in prioritizing 80% of disease states where IVIG can be used. Additional consensus-building rounds will be needed to prioritize the remaining disease states.

PRM221

NETWORK META-ANALYSIS FOR HEALTH TECHNOLOGY SUBMISSIONS WORLDWIDE: A REPORT CHECKLIST FOR NETWORK META ANALYSIS BEST PRACTICES GLOBALLY

Batson S, Greenall G

Abacus International, Oxfordshire, UK

OBJECTIVES: Network meta-analysis (NMA) represents an important and developing method for Health Technology Assessment (HTA). The aim of this study was to review submission guidelines issued by HTA bodies worldwide and produce a checklist for reporting NMA within HTA submissions globally. **METHODS:** The web-based repository of country-specific pharmacoeconomic guidelines maintained by ISPOR was reviewed in January 2015. Guidelines from a number of countries providing sufficient guidance for the use of NMA in HTA submissions were identified and independently reviewed. **RESULTS:** Following review of the available guidance from a number of countries, a single common checklist was developed. The checklist included recommendations relating to five main themes: data; statistical methodology; analyses performed; presentation of results; and technical issues. **CONCLUSIONS:** This reporting checklist provides practical support to health technology manufacturers enabling them to assess the suitability of NMA reports in meeting the requirements of global HTA bodies. In addition, this checklist can be seen as a valid quality tool to critically appraise the reporting of NMAs within HTA.

RESEARCH ON METHODS – Study Design

PRM222

TRANSPARENCY AND REPRODUCIBILITY OF SUPPLEMENTARY SEARCH METHODS IN NICE SINGLE TECHNOLOGY APPRAISAL MANUFACTURER SUBMISSIONS

Witkowski MA, Aldhouse N

Abacus International, Bicester, UK

OBJECTIVES: Systematic reviews (SRs) form an important part of National Institute for Health and Care Excellence (NICE) single technology appraisal (STA) manufacturer submissions. To minimise publication bias when conducting SRs, supplementary searches should be conducted, and should follow the same principles of transparency and reproducibility as database searches. This study aimed to evaluate supplementary search methods used in NICE STA manufacturer submissions. **METHODS:** NICE STAs published between 2011 and 2015 were reviewed. Supplementary search details from manufacturer submissions and related critique from corresponding evidence review group (ERG) reports were extracted. Searches were deemed reproducible if the minimum amount of information required to reproduce searches was reported. **RESULTS:** Of 126 STAs identified, 80 were excluded: appraisal reviews/updates (n=20); appraisal terminated (n=12); no full submission available (n=9); appendices (containing search methods) not published online (n=39). Of 46 included manufacturer submissions, 28 reported conference searches, of which 24 provided enough information for searches to be reproduced. Twenty-one reported clinical trials registry searches, but only seven provided enough information to reproduce these. Thirty-six reported conducting other manual searches, including: manufacturer internal databases (n=24); reference lists (n=20); regulatory body websites (n=11); other websites (n=5); internal experts (n=2). Evidence review groups critiqued omission of supplementary searches in 8 of 18 submissions which lacked searches of conference proceedings, and in 8 of 25 submissions which did not report searching clinical trial registries. The evaluation methods differed between ERGs. **CONCLUSIONS:** Principles of transparency and reproducibility were not fol-

lowed in the majority of manufacturer submissions where supplementary searches were conducted. However, the results from this study are limited due to the low number of appendices published online. Supplementary search methods used in manufacturer submissions should be reported in full and ERGs should be consistent with critique of supplementary search methodology to ensure no evidence is omitted in decision making.

PRM223

INCREASING PRECISION OF REAL-WORLD DATA ESTIMATES: THE IMPORTANCE OF A STEP-WISE PROCESS TO LIMIT DATA COLLECTION ERRORS AND DATA INCOMPLETENESS

Kritikou P[1], de Cock E[2], Stein D[3]
[1]United BioSource Corporation, London, UK, [2]UBC: An Express Scripts Company, London, UK, [3]UBC: An Express Scripts Company, Dorval, QC, Canada

OBJECTIVES: Create a step-wise process to mitigate data collection errors and missing data during all phases of prospective and retrospective observational studies. **METHODS:** Based on three multinational retrospective chart review studies and two multinational time and motion (T&M) studies completed in 2015, key factors were identified during all study phases (design, implementation, conduct, and analysis) that could lead to data collection errors and missing data. For both methodologies, we designed a step-wise process to help identify risk factors and provide effective solutions to improve data quality. **RESULTS:** During study design, study variables should unequivocally be defined with terminology/semantics matching the source document (e.g., medical chart) or what is observed in the real-world. Differences between countries need to be considered. Training using real-time demonstration of electronic data collection (EDC) tool using examples of de-identified patient data is critical for chart reviews. For T&M studies, observers must be trained on accurate data measurement and recording. For a chart review using an EDC tool, logic and edit checks should be built into the EDC tool to limit data errors and incomplete data at entry. For a T&M design, speed of data transmission and fast quality control is essential to allow recall by the data observer. Queries for missing data or outliers should be phrased objectively and clearly. Effectiveness of quality control mechanisms needs to be assessed particularly at the start of data collection, and retraining performed, if needed. **CONCLUSIONS:** Limiting data collection errors and data incompleteness starts at study design. Essential components of a step-wise process include appropriate variable selection and description (terminology/semantics), (re)training, and tailored quality control mechanisms. If such steps are followed, data collected would result in a more accurate dataset, therefore improving the overall quality of study data and precision of study results.

PRM224

A COMPREHENSIVE DISEASE MODEL OF POLYCYSTIC OVARY SYNDROME (PCOS)

Eek D[1], Paty J[2], Black P[3], Celeste Elash CA[3], Reaney M[4]
[1]Astra Zeneca, Molndal, Sweden, [2]Quintiles, Hawthorne, NY, USA, [3]ERT, Pittsburgh, PA, USA, [4]ERT, Peterborough, UK

OBJECTIVES: Polycystic ovary syndrome (PCOS) is one of the most common female endocrine disorders. It presents with a broad range of symptoms and has significant and diverse clinical implications. In order to develop a comprehensive understanding of PCOS, a (conceptual) disease model was developed. **METHODS:** The disease model was generated based on three lines of concept evaluation research; (1) a targeted literature review, (2) interviews with clinical experts; (3) concept elicitation interviews with patients, for which data was recorded, transcribed and coded. Collectively, this provided a comprehensive list of the sign, symptom and impact concepts most important and relevant to women with PCOS. **RESULTS:** Nineteen peer-reviewed articles were included in the literature review. Five clinical experts (USA, Turkey, Netherlands) and 20 PCOS patients (mean (SD) age 29.2 (5.9) years) were included in 1:1 qualitative interviews. Concept saturation was observed in patient interviews. Significant overlap was seen in the sign, symptom and impact concepts of PCOS across the three lines of evidence. Signs/symptoms were categorized into pain, infertility, hirsutism, alopecia, acne, menstruation (e.g. irregular menstruation, heavy bleeding), bloating, weight-related (e.g. weight gain, fluctuations), and metabolic abnormality (i.e., obesity, difficulty with weight loss, etc) symptoms. Some symptoms, such as pain at non-menstrual times, were uniquely reported by patients. Impacts of PCOS included sleep disturbance, emotional functioning, social role functioning and physical functioning. Compensatory behaviours (e.g. hair removal, diet changes, use of medication) were common. The relationship between these concepts is presented in a disease model. **CONCLUSIONS:** This is the first known comprehensive disease model for PCOS. It shows many of the defining features of the condition can only be accurately and reliably captured by asking patients how they feel and function. This work underscores the need for measurement of PCOS from the patient perspective using a patient reported outcome (PRO).

PRM225

EVALUATION OF QUALITY ASSESSMENT TOOLS FOR NON-RANDOMISED CONTROLLED TRIALS ASSESSING SURGICAL INTERVENTIONS: A SYSTEMATIC REVIEW OF SYSTEMATIC REVIEWS

Nurmatov UB, Xiong T, Kroes MA
Abacus International, Bicester, UK

OBJECTIVES: Evaluating the effectiveness and safety of surgical techniques, operations, and involved medical devices is relatively difficult, sometimes unethical, with randomised controlled trials (RCTs). Non-randomised designs are commonly applied and used to inform decision making. Quality assessment (QA) methods for these studies have previously been reviewed, but not specifically for their applicability to non-RCTs in surgical interventions. The objectives of this systematic review were to evaluate which QA tools have been used in this research field and to critically appraise these tools. **METHODS:** We systematically searched three electronic databases (MEDLINE, Embase and Cochrane Library) and Health Technology Assessments. Systematic reviews appraising the quality of non-RCTs on surgical interventions were included. **RESULTS:** In total, 1,741 potentially relevant citations were identified. After removing duplicates, 1,525 citations were screened. Of these, 159 full text references were reviewed and 85 systematic reviews met predefined inclusion criteria. Five QA methods were most commonly employed: Newcastle-Ottawa Quality Assessment Scale (NOS) or modified NOS (28%); checklists developed by authors (15%); the Cochrane checklist or modified version (11%); modified checklists from other authors (5%); applying disease-specific QA tools (5%). The reliability and applicability of the most commonly employed tool in this research field, NOS, were questioned in included reviews, corresponding with concerns on the validity of the NOS reported in recently published literature of research methodology. **CONCLUSIONS:** The available evidence demonstrates a lack of consensus on the use of QA tools for non-RCTs assessing surgical interventions. Various methods have been adapted or newly developed by researchers, and the most commonly applied QA tool (NOS) may not be fit for purpose in this field of research. There is an urgent need for a validated QA tool to appraise the quality of evidence to help inform evidence-based decision making on the use of surgical devices and types of surgical approaches.

PRM226

INCIDENCE AND PREVALENCE ESTIMATIONS BASED ON CLAIMS DATA – NEW METHODOLOGICAL CONSIDERATIONS

Prosser C[1], Altevers J[1], Hickstein L[2], Walker J[2], Braun S[1]
[1]Xcenda GmbH, Hannover, Germany, [2]Elsevier Health Analytics, Health Risk Institute, Berlin, Germany

OBJECTIVES: Scientific analyses with claims data such as burden of disease analyses are often based on incidence and prevalence estimates. Latest methodological considerations indicate that the diagnosis-free observation period should be extended as much as possible to not overestimate the incidence in chronic diseases. Aim of this study was to evaluate the impact of expanding the diagnosis timeframe for the incidence as well as the prevalence estimates. **METHODS:** This methodological analysis focused on the chronic diseases diabetes mellitus (DM) and multiple sclerosis (MS) in 2013 in Germany and was based on anonymized data from the Health Risk Institute Research Database. Patients continuously insured for six years (2008-2013) were included in the study (n=3,026,154). Incidence changes due to different diagnosis-free intervals before a diagnosis in 2013 (1 to 5 years) were assessed. Correspondingly, the prevalence estimation for 2013 was varied by expanding the timeframe for diagnosis from 1 year up to 5 years, as it was assumed that chronic diseases identified in previous years persist until 2013. Moreover, disease-specific validation of the diagnosis codes was applied as sensitivity analysis. **RESULTS:** DM incidence was 24% higher when a 1-year diagnosis-free observation period was applied compared to 5 years (25% in MS). When expanding the prevalence timeframe up to 5 years, the prevalence estimation increased by 14% in DM and 21% in MS, respectively. The relative proportion of incidence to prevalence also changed by varying the utilized timeframe. Out of the prevalent diabetes patients in 2013 10.1% were incident when a 1-year timeframe was considered, whereas 7.3% were incident when 5 years were applied (11.7% and 7.9% in MS, respectively). **CONCLUSIONS:** The methodological concepts should coincide when estimating both the incidence and the prevalence of chronic diseases in claims data. Estimates may be biased especially when only short timeframes are utilized.

PRM227

STATISTICAL ASSESSMENT OF A CASE-FINDING ALGORITHM FOR IDENTIFYING NON-SMALL CELL LUNG CANCER (NSCLC) PATIENTS IN ADMINISTRATIVE CLAIMS DATABASES

Turner RM[1], Croft E[2], Chen Y[1], Tunceli O[3], Elesinmogun C[2], Fernandes A[2]
[1]HealthCore, Wilmington, DE, USA, [2]AstraZeneca, Gaithersburg, MD, USA, [3]HealthCore Inc., Wilmington, DE, USA

OBJECTIVES: The ICD-9-CM coding system does not differentiate between small cell lung cancers (SCLC) and NSCLC, which poses a challenge for database research on forms of lung cancer. We examined the accuracy of an algorithm designed to identify likely NSCLC cases among lung cancer patients in a claims-database. **METHODS:** Lung cancer patients were selected from the HealthCore Integrated Research Environment (HIRE)-Oncology database which combines US administrative claims database, and the clinical oncology data (type, stage, etc.) on lung cancer patients. Index event was defined as the patient's first lung cancer diagnosis during 6/1/14 to 1/31/15 in the claims database. Eligibility criteria were; ≥1 lung cancer diagnosis & > 12 months continuous pre-index enrolment in the claims database; and presence in the oncology database. A treatment regimen algorithm was used to identify NSCLC patients from claims data. This was assessed against the cancer type information from the oncology database. Diagnostic accuracy of the algorithm was assessed using statistical measures; Sensitivity, Specificity, False Positive Fraction (FPF), Positive Predictive Validity (PPV), Negative Predictive Validity (NPV), Positive Likelihood Ratio (LR+), Negative Likelihood Ratio (LR-), Diagnostic Odds Ratio (DOR), and Agreement (kappa:k). **RESULTS:** 585 lung cancer patients (mean age = 62, 53% male) met all eligibility criteria for analysis. The algorithm classified 464 (79%) patients as NSCLC and 121 (21%) as SCLC; whereas, the clinical data classified 513 (88%) patients as NSCLC and 72 (12%) as SCLC. Algorithm sensitivity was 86% and specificity was 71%. The FPF = .29%, PPV= 96%, and NPV= 42%. LR+ = 2.96; LR- = 0.19; and DOR= 15.37. Chance adjusted k = .60. **CONCLUSIONS:** The algorithm showed good statistical properties for identifying NSCLC patients in claims data except for a high false positive fraction. Future research should focus on improving the algorithm's specificity.

PRM228

SYSTEMATIC LITERATURE REVIEW OF ADJUNCTIVE ANTI-EPILEPTIC DRUG TRIALS IN PATIENTS WITH PRIMARY GENERALIZED TONIC-CLONIC SEIZURES ILLUSTRATES CHANGES IN STANDARD OF CARE OVER 12-20 YEARS

Tsong WH[1], Gomez-Ulloa D[2], Li H[3], Forsythe A[1], Kramer L[1]
[1]Eisai Inc., Woodcliff Lake, NJ, USA, [2]IMS Health, Barcelona, Spain, [3]IMS Health, Alexandria, VA, USA